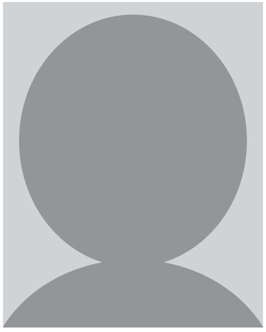
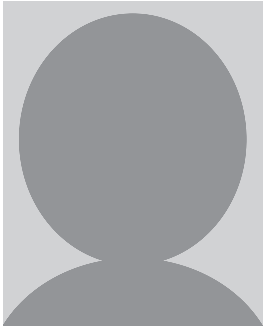
問題解決の数理（’１７）

- 収録本番とは多少異なっていることがあります.
- 内容の間違いのご指摘は歓迎します.
- 「完全に無保証」です.

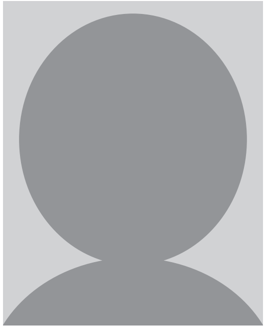


- 今回の講義では，非線形最適化法についてお話します.
- 線形最適化問題は，目的関数と制約がすべて線形式で記述されています.
- そうすると，目的関数ないしは制約に線形でない式が含まれると，非線形最適化問題になるということが想像できると思いますが，実際，その通りです.

非線形最適化問題



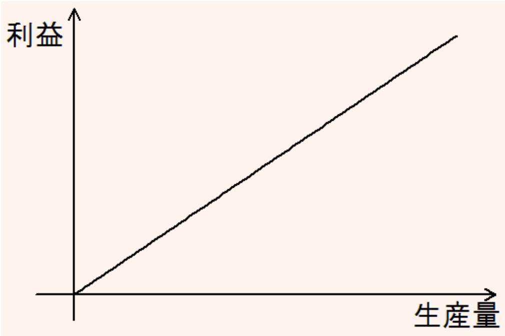
- 現実世界に非線形問題が多数あります…
- むしろ，厳密に線形性が成り立つことが例外，と言った方が正確かもしれません．
- 例えば，代表的な線形最適化問題である，生産計画問題では，例えば「製品一キログラムあたり一万円の利益の見込み」としています．



- そうしますと，百キログラム生産すれば百万円の利益見込み，一万キログラム生産すれば一億円の利益見込みということになります．
- グラフにすると次のようになります．

非線形最適化問題

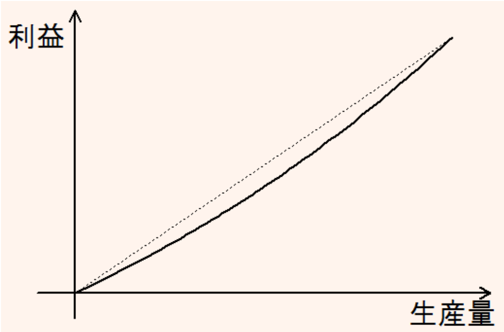
生産量と利益見込み (線形関係)



- 実際の生産では，大量生産をすると，単位生産量当たりのコストが下がることが多いです．
- そうしますと，もし，同じ値段で販売できれば，（ゆっくり）利益見込みは次のグラフに示しますように...

非線形最適化問題

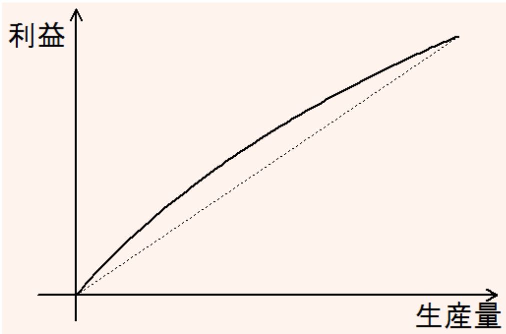
生産量と利益見込み (非線形関係)



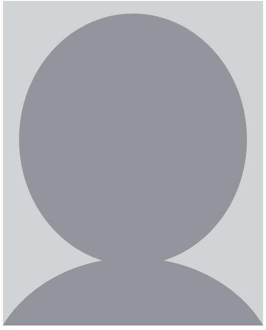
- 下に凸の上昇カーブを描きます.
- 逆に生産量が多くなりすぎますと，値下げをしなければ売れなくなり，（ゆっくり）利益見込みは次のグラフに示しますように...

非線形最適化問題

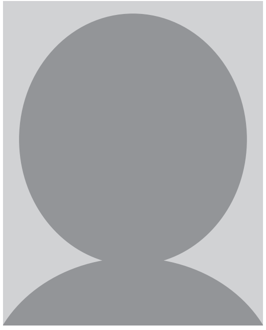
生産量と利益見込み (非線形関係)



- 上に凸の上昇カーブを描きます．
- もちろん，線形近似が成り立つ問題は線形最適化法で効率的に解いて何も問題はありません．
- しかし，線形近似が成り立たない問題は，非線形問題として解く必要があり，そのような問題は多々あります．



- 非線形と聞くと，難しそうな印象を与えるかもしれませんが，問題の定式化自体は非線形だから難しいということはありません．
- しかし，問題を解く計算の手続きは複雑になり，一回の講義ではとても説明し切れません．
- 実際の計算は計算機ソフトウェアを利用して行うのが一般的ですので，計算に関してはあまり深入りせず，制約のない非線形最適化法についてのみ説明します．



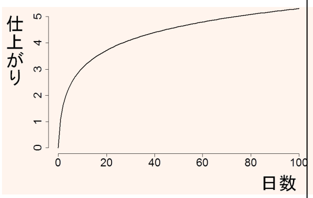
- 制約なしの非線形最適化法は，制約のある非線形最適化法よりかなり容易です．
- 制約なしの非線形最適化問題も多数ありますし，制約のある非線形最適化法を学習する上でも役に立ちますので，そのつもりで聞いて下さい．
- それでは，計算の説明の前に，制約のある非線形最適化問題と制約なしの非線形最適化問題の例を示します．
- まずは，制約のある非線形最適化問題の例です．

非線形最適化問題

制約のある非線形最適化問題

- 工程 i $(i = 1, 2, \cdots, n)$ に x(日) かけた時の仕上がりの程度

$$\log(a_i x + 1) \quad a_i > 0$$



- 各工程は最低 $t_i (> 0)$ 日かかる
- 全工程を T 日以内に終わらせなければならない
- 仕上がりの和を最大にするには，各工程に何日かければよいか

- 次のような問題を考えます．
- 会社員の A さんは自分自身で自宅を建てようとしています．
- 各工程に時間をかければその分仕上がりは良くなりますが，仕上がりの程度は時間に比例するわけでなく，図に示すような上に凸の曲線になります．

- 工程数は全部で n であり，工程 i の仕上がりの程度は，その工程にかけた時間を x (日) としたとき，

$$\log(a_i x + 1)$$

です．
- ここで，各 a_i は正の値をとるとします．

- また，各工程は最低 $t_i (> 0)$ 日かかります．
- 全工程をラージ T 日以内に終わらせなければならないとき，仕上がり程度の総和を最大にするには，各工程に何日かければよいか…という問題です．
- それでは，この問題を非線形最適化問題として定式化します．

非線形最適化問題

制約のある非線形最適化問題

- 工程 i に x_i 日かける
- 目的関数 z は各工程の仕上がり程度の総和

$$z = \sum_{i=1}^{n} \log(a_i x_i + 1)$$

- 工程 i に x_i 日かけるとします.
- 目的関数 z は各工程の仕上がり程度の総和ですから，仕上がり程度 $\log(a_i x_i + 1)$ の総和となります.
- 目的関数 z を最大にする各 x_i を決定する問題です.
- 次に制約条件を定式化します.

非線形最適化問題

制約のある非線形最適化問題

- 工程 i には最低 t_i(日) かかる

$$x_i \geq t_i \ \ (i = 1, 2, \cdots n)$$

- 全工程を T 日以内に終わらせる

$$\sum_{i=1}^{n} x_i \leq T$$

- 工程 i には最低 t_i(日) かかりますから，

$$x_i \geq t_i$$

  となります．

- 全工程をラージ T 日以内に終わらせる必要がありますから，x_i の総和はラージ T 以下となります．

- 以上をまとめますと，この問題は次のように非線形最適化問題として定式化されます．

非線形最適化問題

制約のある非線形最適化問題

$$
\begin{array}{rl}
\text{最大化} & z = \sum_{i=1}^{n} \log(a_i x_i + 1) \\
\text{制約条件} & x_i \geq t_i \ \ (i = 1, 2, \cdots, n) \\
 & \sum_{i=1}^{n} x_i \leq T
\end{array}
$$

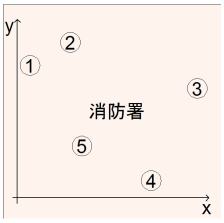
- この定式化では，目的関数が非線形になっている点で線形最適化問題とは異なります．
- 制約条件には非線形の式はありませんが，非線形最適化問題一般では，制約条件にも非線形の式が入ることもあります．

- このように，非線形最適化問題の定式化は，非線形の式が入るだけであり，非線形だから難しいということはありません．
- 今度は，制約のない非線形最適化問題の例です．

非線形最適化問題

制約のない非線形最適化問題

- 5 つの人口密集地区 (1, 2, 3, 4, 5)
- 消防署はこれら 5 つの地区への直線距離の和が最小になる位置に設置
- 消防署を設置すべき位置を求めよ



地区	x	y
1	1	12
2	4	14
3	15	10
4	11	2
5	5	5

- 次のような問題を考えます．
- B 市では近年の人口増加にともない消防署を新設することになりました．
- 市内には図に示すような 5 つの人口密集地区 (1, 2, 3, 4, 5) があります．
- 消防署はこれら 5 つの地区への直線距離の和が最小になる位置に設置することにします．
- 消防署を設置すべき位置の座標を求めよ…という問題です．
- 表には各地区の位置の座標が示してあります．
- もちろん地区は点ではありませんので，ここで示す座標は，重心のような適当な指標であると考えて下さい．
- それでは，この問題を定式化します．

非線形最適化問題

制約のない非線形最適化問題

- 消防署の位置の座標を (x, y)
- 地区 i $(i = 1, 2, 3, 4, 5)$ の位置の座標　(x_i, y_i)
- 消防署と地区 i の直線距離は，

$$\sqrt{(x - x_i)^2 + (y - y_i)^2}$$

- 消防署の位置の座標を (x, y) とします．
- 地区 i の位置の座標を (x_i, y_i) としますと，消防署と地区 i の直線距離は，$x - x_i$ の 2 乗と $y - y_i$ の 2 乗の和の平方根となります．

非線形最適化問題

制約のない非線形最適化問題

- 消防署の位置の座標を (x, y)
- 地区 i $(i = 1, 2, 3, 4, 5)$ の位置の座標を (x_i, y_i)
- 目的関数 z は消防署と各地区の直線距離の和で

$$z = \sum_{i=1}^{5} \sqrt{(x - x_i)^2 + (y - y_i)^2}$$

- z を最小化する x, y を求める問題

- 目的関数 z は消防署と各地区の直線距離の総和ですから，この式のようになります．
- この z を最小化する x と y を決定する問題です．

- 線形最適化問題では，制約条件がないと，目的関数を無限に大きく，あるいは小さくすることができますので，制約がない問題は考えられませんが，非線形最適化問題では，この例のように，制約がなくても目的関数の最適値，最適解が存在する問題も存在します．

非線形最適化問題

非線形最適化問題の一般形

$$
\begin{array}{rll}
\text{最小化} & f(\boldsymbol{x}) & \\
\text{制約} & g_i(\boldsymbol{x}) = 0 & (i = 1, 2, \cdots, m) \\
 & h_j(\boldsymbol{x}) \leq 0 & (j = 1, 2, \cdots, l)
\end{array}
$$

- $\boldsymbol{x}$ は n 変数のベクトル $[x_1, x_2, \cdots, x_n]^T$

- 非線形最適化問題の一般形は，ここに示すようになります.
- 目的関数および制約式の一方，または両方に非線形の式を含みます.
- 線形最適化問題と同様に，制約は等式制約も不等式制約も含みます.

非線形最適化問題

今回の前提

- 制約のない非線形最適化問題の最適化
- 問題は最小化問題
- 目的関数 $f(\boldsymbol{x})$ は考慮すべき $\boldsymbol{x}$ の領域において 2 階微分可能
- 2 階微分は連続関数

- 今回は，制約のない非線形最適化問題の最適化法についてお話します．
- 1 変数関数の最適化については，微分法の応用として，極値，すなわち極小値や極大値を求めたり，最小値，最大値を求めることを学習したことがある方も少なくないと思います．
- ここでは，1 変数関数の最適化と多変数関数の最適化を対応付けながら，制約なしの非線形最適化問題における最適性の条件について説明します．
- 以降の説明を円滑にするため，幾つかの前提を設けます．
- まず，問題は最小化問題とします．すなわち，最適化と言った場合，それは最小化を意味します．
- それから，目的関数 $f(\boldsymbol{x})$ は考慮すべき $\boldsymbol{x}$ の領域において 2 階微分可能で，2 階微分は連続関数であると仮定します．

非線形最適化問題

最適性の条件

- 任意の $\boldsymbol{x}$ に対して $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ を満たす $\boldsymbol{x}^*$
  $\cdots$ 関数 f の最小化問題の大域最適解
- $\boldsymbol{x}^*$ の周囲の $\boldsymbol{x}$ に対して $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ を満たす $\boldsymbol{x}^*$
  $\cdots$ 関数 f の最小化問題の局所最適解
- 大域最適解は局所最適解であるが，局所最適解は一般には大域最適解とはならない

- 任意の $\boldsymbol{x}$，ここで $\boldsymbol{x}$ は n 個の変数を並べたベクトルを表わしますが，任意の $\boldsymbol{x}$ に対して，関数 $f(\boldsymbol{x})$ の値が $f(\boldsymbol{x}^*)$ より小さくならない時，$\boldsymbol{x}^*$ は最小化問題の大域最適解となります．
- このとき，$f(\boldsymbol{x}^*)$ は関数 $f(\boldsymbol{x})$ の最小値になっているということです．
- 一方，$\boldsymbol{x}^*$ の周囲の $\boldsymbol{x}$ に対して，関数 $f(\boldsymbol{x})$ の値が $f(\boldsymbol{x}^*)$ より小さくならない時，$\boldsymbol{x}^*$ は最小化問題の局所最適解となります．
- このとき，$f(\boldsymbol{x}^*)$ は関数 $f(\boldsymbol{x})$ の極小値になっているということです．
- つまり，$\boldsymbol{x}^*$ から離れたところまで探索すると，$f(\boldsymbol{x}^*)$ より $f(\boldsymbol{x})$ が小さくなるような $\boldsymbol{x}$ が存在する可能性があります．
- 言い換えますと，大域最適解は局所最適解ですが，局所最適解は一般には必ずしも大域最適解とはなりません．

非線形最適化問題

- 1 変数関数 $f(x)$
- 関数 $f(x)$ の増減を調べるには，関数の変化率を表す (1 階) 導関数 (微分)

$$f'(x) = \frac{df(x)}{dx} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

- 導関数の変化率を表す 2 階導関数

$$f''(x) = \frac{d^2 f(x)}{dx^2} = \lim_{\Delta x \to 0} \frac{f'(x+\Delta x) - f'(x)}{\Delta x}$$

- まず，1 変数関数の最適性の条件について説明します．
- 先ほどお話しましたように，1 階微分を利用します．
- 1 変数関数 $f(x)$ の増減を調べるには，関数の変化率を表す 1 階の導関数，1 階微分を利用します．
- 1 階の導関数はここに示す式により定義されます．
- 1 階の導関数は関数 $f(x)$ の x における変化率を示しています．
- 1 階の導関数の導関数は 2 階の導関数，2 階微分と呼ばれます．

非線形最適化問題

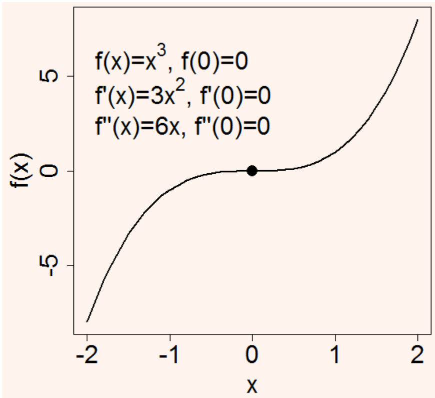
最適性の条件

- $f'(x)$ は x における $f(x)$ の変化率 (接線の傾き)
- x において，$f'(x)$ が正なら x において $f(x)$ は増加
- $f'(x)$ が負なら x において $f(x)$ は減少
- x において $f(x)$ が極小値あるいは極大値
$\rightarrow$ x において $f'(x) = 0$
- $f'(x) = 0$ なら必ずしも極小値あるいは極大値
になっている訳ではない

- 関数 $f(x)$ の 1 階導関数 $f'(x)$ は，x における $f(x)$ の変化率を示しています．
- 幾何的には x における $f(x)$ の接線の傾きになります．
- x において，$f'(x)$ が正なら x において $f(x)$ は増加していて，$f'(x)$ が負なら x において $f(x)$ は減少しています．
- x において $f(x)$ が極小値あるいは極大値になっているなら，x において $f'(x)$ は 0 となります．
- 逆は必ずしも成り立たず，$f'(x) = 0$ でも $f(x)$ が極小値にも極大値にもならない場合があります．

非線形最適化問題

最適性の条件



- 例えば，$f(x) = x^3$ という関数を考えます．
- 1 階微分は $f'(x) = 3x^2$ で，$x = 0$ において，$f'(x)$ は 0 になります．
- しかし，図から明らかなように $x = 0$ は $f(x)$ の極値ではありません．
- なお，2 階微分は $f''(x) = 6x$ で，$x = 0$ において，$f''(x)$ は 0 になります．

非線形最適化問題

最適性の条件

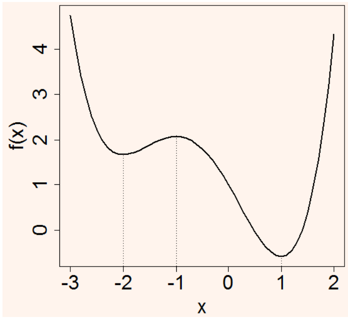
- $f'(x) = 0$ のとき，x において $f(x)$ が極小値であるか極大値であるかの判別
- $f''(x)$ が正なら $f(x)$ は下に凸な曲線 ($\smile$) $\cdots$ 極小値
- $f''(x)$ が負なら $f(x)$ は上に凸な曲線 ($\frown$) $\cdots$ 極大値
- $f''(x)$ が 0 の場合，これだけからでは最適性の判定はできない

- $f'(x) = 0$ のとき，x において $f(x)$ が極小値であるか，極大値であるかの判別は，次のように行います．
- $f''(x)$ が正なら $f(x)$ は下に凸な曲線になりますので，$f(x)$ は極小値になります．
- $f''(x)$ が負なら $f(x)$ は上に凸な曲線になりますので，$f(x)$ は極大値になります．
- $f''(x)$ が 0 の場合ですが，この場合は残念ながら，これらの情報だけからでは，最適性の判定はできません．
- それでは，例を見てみましょう．

非線形最適化問題

最適性の条件

$$f(x) = \frac{1}{4}x^4 + \frac{2}{3}x^3 - \frac{1}{2}x^2 - 2x + 1$$



- 4 次関数
$$f(x) = \frac{1}{4}x^4 + \frac{2}{3}x^3 - \frac{1}{2}x^2 - 2x + 1$$
を考えます.
- この関数のグラフは図に示す通りです.

非線形最適化問題

最適性の条件

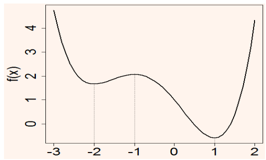
$$
\begin{aligned}
f(x) &= \frac{1}{4}x^4 + \frac{2}{3}x^3 - \frac{1}{2}x^2 - 2x + 1 \\
f'(x) &= x^3 + 2x^2 - x - 2 = (x+2)(x+1)(x-1) \\
\bullet\ x &= -2,\ -1,\ 1\ \rightarrow\ f'(x) = 0 \\
f''(x) &= 3x^2 + 4x - 1 \\
\bullet\ x &= \frac{-2 \pm \sqrt{7}}{3} \rightarrow f''(x) = 0
\end{aligned}
$$

- $f(x)$ の 1 階および 2 階の導関数は，これらの式になります．
- 1 階の導関数 $f'(x)$ が 0 になるのは，$x = -2, -1, 1$ の時です．
- 2 階の導関数 $f''(x)$ が 0 になるのは，$x = \frac{-2\pm\sqrt{7}}{3}$ の時です．

非線形最適化問題

最適性の条件

$$f(x) = \frac{1}{4}x^4 + \frac{2}{3}x^3 - \frac{1}{2}x^2 - 2x + 1$$



x	< -2	-2		$\frac{-2-\sqrt{7}}{3}$		-1		$\frac{-2+\sqrt{7}}{3}$		1	> 1
$f'(x)$	$-$	0	$+$	$+$	$+$	0	$-$	$-$	$-$	0	$+$
$f''(x)$	$+$	$+$	$+$	0	$-$	$-$	$-$	0	$+$	$+$	$+$
$f(x)$	$\smile$	$\frac{5}{3}$ 極小	$\smile$		$\frown$	$\frac{25}{12}$ 極大	$\frown$		$\smile$	$\frac{-7}{12}$ 極小	$\smile$

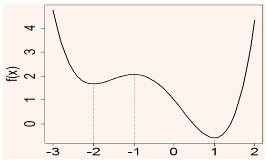
- これらを基に増減表を作成すると，この表のようになります．
- x が -2 より小さい時は，1 階導関数 $f'(x)$ が負で，$f(x)$ は減少しています．
- 2 階導関数 $f''(x)$ は正なので，$f(x)$ は下に凸で，減少の割合は小さくなっています．
- x が -2 となると，$f'(x)$ が 0 となります．
- $f''(x)$ は正なので，$f(x)$ は $x = -2$ において極小値となります．
- x が -2 より大きくとなると，$f'(x)$ は正になりますので，$f(x)$ は増加し，また $f''(x)$ が正になりますので，$f(x)$ は下に凸で，増加の割合も大きくなります．
- x が $\frac{-2-\sqrt{7}}{3}$ になると，$f''(x)$ は 0 になり，x が $\frac{-2-\sqrt{7}}{3}$ より大きくなると，$f''(x)$ は負になります．
- この間も $f'(x)$ は正ですので，$f(x)$ は増加を続けていますが，x が $\frac{-2-\sqrt{7}}{3}$ より大きくなると，$f(x)$ は上に凸になり，増加の割合は小さくなります．
- そして，x が -1 になると，$f'(x)$ は 0 になります．
- この時，$f''(x)$ は負ですから，$f(x)$ は $x = -1$ で極大値となります．
- x が -1 より大きくなると $f'(x)$ は負になり，$f(x)$ は減少します．
- この時，$f''(x)$ は負ですから，$f(x)$ は上に凸で，減少の割合は大きくなります．
- x が $\frac{-2+\sqrt{7}}{3}$ になると，$f''(x)$ は 0 になり，x が $\frac{-2+\sqrt{7}}{3}$ より大きくなると，$f''(x)$ は正になります．

- この間も $f'(x)$ は負ですので，$f(x)$ は減少を続けていますが，x が $\frac{-2+\sqrt{7}}{3}$ より大きくなると，$f(x)$ は下に凸になり，減少の割合は小さくなります．
- そして，x が 1 になると，$f'(x)$ は 0 になります．
- この時，$f''(x)$ は正ですから，$f(x)$ は $x = 1$ で極小値となります．
- x が 1 より大きくなると，$f'(x)$ は正となり，$f''(x)$ は引き続き正です．
- したがって，$f(x)$ は増加し，増加の割合も大きくなります．

非線形最適化問題

最適性の条件

$$f(x) = \frac{1}{4}x^4 + \frac{2}{3}x^3 - \frac{1}{2}x^2 - 2x + 1$$



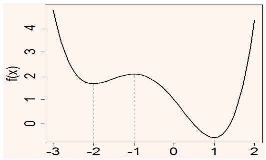
x	< -2	-2		$\frac{-2-\sqrt{7}}{3}$		-1		$\frac{-2+\sqrt{7}}{3}$		1	> 1
$f'(x)$	$-$	0	$+$	$+$	$+$	0	$-$	$-$	$-$	0	$+$
$f''(x)$	$+$	$+$	$+$	0	$-$	$-$	$-$	0	$+$	$+$	$+$
$f(x)$	$\smile$	$\frac{5}{3}$ 極小	$\smile$		$\frown$	$\frac{25}{12}$ 極大	$\frown$		$\smile$	$\frac{-7}{12}$ 極小	$\smile$

- こうして $f(x)$ の増減を調べました．
- その結果，$x = -2$ および $x = 1$ において，$f(x)$ は極小値となることが分かりました．
- これらの時の $f(x)$ を比較すると…

非線形最適化問題

最適性の条件

$$f(x) = \frac{1}{4}x^4 + \frac{2}{3}x^3 - \frac{1}{2}x^2 - 2x + 1$$



x	< -2	-2		$\frac{-2-\sqrt{7}}{3}$		-1		$\frac{-2+\sqrt{7}}{3}$		1	> 1
$f'(x)$	$-$	0	$+$	$+$	$+$	0	$-$	$-$	$-$	0	$+$
$f''(x)$	$+$	$+$	$+$	0	$-$	$-$	$-$	0	$+$	$+$	$+$
$f(x)$	$\smile$	$\frac{5}{3}$ 極小	$\smile$		$\frown$	$\frac{25}{12}$ 極大	$\frown$		$\smile$	$\frac{-7}{12}$ 最小	$\smile$

- $x = -2$ のときに $f(x) = 5/3$，$x = 1$ のときに $f(x)$ は $-\frac{7}{12}$ であり，$x = 1$ のときに最小値 $-\frac{7}{12}$ になることが分かります．

非線形最適化問題

最適性の条件

- n 変数関数 $f(\boldsymbol{x})$ $(\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^T)$ の偏導関数

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i} = \lim_{\Delta x_i \to 0} \frac{f(x_1, \cdots, x_i + \Delta x_i, x_{i+1}, \cdots, x_n) - f(x_1, \cdots, x_i, x_{i+1}, \cdots, x_n)}{\Delta x_i}$$

$$\frac{f(x_1, \cdots, x_i + \Delta x_i, x_{i+1}, \cdots, x_n) - f(x_1, \cdots, x_i, x_{i+1}, \cdots, x_n)}{\Delta x_i}$$

- x_i のみを独立変数と見なして微分

- 今度は，多変数関数の最適性の条件について説明します．
- まず，1 変数関数における，1 階および 2 階の導関数の概念を多変数関数に拡張します．
- n 変数のベクトル $\boldsymbol{x}$ の関数 $f(\boldsymbol{x})$ について考えます．
- $f(\boldsymbol{x})$ の x_i による 1 階の偏導関数，1 階偏微分はここに示す式になります．
- n 個の変数のうち，x_i 以外の変数は固定して，x_i のみが変数の関数と見なして，微分するというものです．

非線形最適化問題

最適性の条件

- 偏導関数 $\frac{\partial f(\boldsymbol{x})}{\partial x_i}$， $f_{x_i}(\boldsymbol{x})$ を成分とする n 次元ベクトル

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$$

勾配ベクトル

- $f(\boldsymbol{x})$ の x_i による 1 階の偏導関数は…このように表記しますが，1 階の偏導関数を並べたベクトルを勾配ベクトルと呼びます．
- 逆三角形の記号は「ナブラ演算子」[1]と呼ばれ，関数 f を変数ベクトル $\boldsymbol{x}$ の各成分で偏微分したものを並べたベクトルを出力する演算子です．

[1] テロップ

非線形最適化問題

最適性の条件

- 2 階の偏導関数 $\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}$, $f_{x_i x_j}(\boldsymbol{x})$ を成分とする $n \times n$ 行列

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} \\ & & \vdots & \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n^2} \end{bmatrix}$$

ヘッセ行列

- 関数 $f(\boldsymbol{x})$ の $\boldsymbol{x}$ における，変数 x_j による 1 階の偏導関数をさらに x_i により偏微分した関数は，$f(\boldsymbol{x})$ の 2 階の偏導関数になり，このように表記します.
- この 2 階の偏導関数を i, j 成分とする $n \times n$ 行列をヘッセ行列と呼びます.

非線形最適化問題

最適性の条件

$n \times n$ 行列 $\boldsymbol{M}$

- 任意の n 次元ベクトル $\boldsymbol{x}$ に対して

$$\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} \geq 0$$

を満たすとき，行列 $\boldsymbol{M}$ は半正定値

- 任意の実数 x に対して

$$mx^2 \geq 0$$

を満たすとき，$m \geq 0$

- $n \times n$ 行列 $\boldsymbol{M}$ を考えます．
- 任意の n 次元ベクトル $\boldsymbol{x}$ に対して，この式を満たす時，行列 $\boldsymbol{M}$ は半正定値であると言います．
- これは，実数 m に関して，任意の実数 x に対して

$$mx^2 \geq 0$$

を満たすとき，$m \geq 0$ となることに対応します．
- つまり，行列 $\boldsymbol{M}$ が半正定値であることは，実数 m が非負であることに相当します．

非線形最適化問題

最適性の条件

$n \times n$ 行列 $\boldsymbol{M}$
- 零ベクトルを除く任意の n 次元ベクトル $\boldsymbol{x}$ に対して

$$\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} > 0$$

を満たすとき，行列 $\boldsymbol{M}$ は正定値
- 0 を除く任意の実数 x に対して

$$mx^2 > 0$$

を満たすとき，$m > 0$

- $n \times n$ 行列 $\boldsymbol{M}$ に関して，
- 零ベクトルを除く任意の n 次元ベクトル $\boldsymbol{x}$ に対して，この式を満たす時，行列 $\boldsymbol{M}$ は正定値であると言います．
- これは，実数 m に関して，0 を除く任意の実数 x に対して

$$mx^2 > 0$$

を満たすとき，$m > 0$ となることに対応します．
- つまり，行列 $\boldsymbol{M}$ が正定値であることは，実数 m が正であることに相当します．

非線形最適化問題

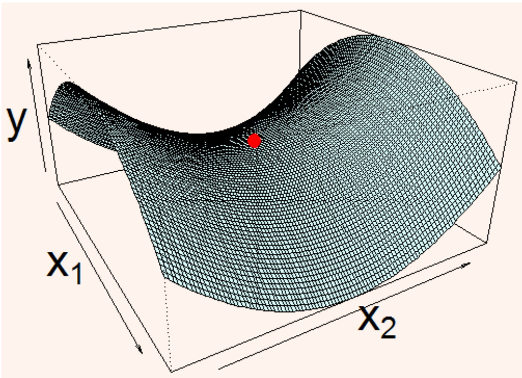
最適性の条件

- $\boldsymbol{x}$ が局所最適解であれば
$$\nabla f(\boldsymbol{x}) = \mathbf{0}$$
- $\boldsymbol{x}$ が局所最適解であるための必要条件
$\cdots$ 1 次の必要条件
- $\nabla f(\boldsymbol{x}) = \mathbf{0}$ を満たす $\boldsymbol{x}$
$\cdots$ $f(\boldsymbol{x})$ の停留点

- $\boldsymbol{x}$ が関数 $f(\boldsymbol{x})$ の局所最適解であれば，$\boldsymbol{x}$ における勾配ベクトルは零ベクトルになります．すなわち，
$$\nabla f(\boldsymbol{x}) = \mathbf{0}$$
となります．
- 勾配ベクトルが零ベクトルであることは，$\boldsymbol{x}$ が局所最適解であるための必要条件ですので，この条件は 1 次の必要条件と呼ばれます．
- $\boldsymbol{x}$ において勾配ベクトルが零ベクトルになる時，$\boldsymbol{x}$ は $f(\boldsymbol{x})$ の停留点と呼ばれます．
- これは 1 変数関数において，関数 $f(x)$ が極値である時，$f'(x) = 0$ であることに相当します．
- 逆は必ずしも成り立たず，勾配ベクトルが零ベクトルであっても局所最適解にならないことがあります．

非線形最適化問題

最適性の条件



- 例えば，この図で 2 変数関数 y は赤い点において，勾配ベクトルが零ベクトルになりますが，局所最適解ではありません．
- この点では，x_2 に関しては谷底になっていますが，x_1 に関しては山頂になっています．
- このような点は，曲面の形状が，馬に乗る時に使う鞍に似ていることから，鞍の点と書いて，鞍点といいます．

非線形最適化問題

最適性の条件

- $\boldsymbol{x}$ が局所最適解であれば

$$\nabla f(\boldsymbol{x}) = \mathbf{0}$$

かつ

$$\nabla^2 f(\boldsymbol{x}) \text{ は半正定値}$$

- $\boldsymbol{x}$ が局所最適解であるための必要条件
  $\cdots$ 2 次の必要条件

- 関数 $f(\boldsymbol{x})$ が $\boldsymbol{x}$ において極小になる時，勾配ベクトルが零ベクトルになることに加えて，ヘッセ行列が半正定値になります．
- これは，最適性の 2 次の必要条件と呼ばれます．

非線形最適化問題

最適性の条件

$$\nabla f(\boldsymbol{x}) = \boldsymbol{0}$$

かつ

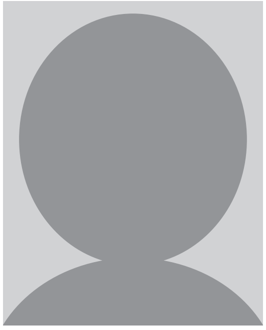
$$\nabla^2 f(\boldsymbol{x}) \text{ は正定値}$$

であれば

- $\boldsymbol{x}$ は局所最適解
- $\boldsymbol{x}$ が局所最適解であるための十分条件
  $\cdots$ 2 次の十分条件

- 勾配ベクトルが零ベクトルになることに加えて，ヘッセ行列が正定値であれば，関数 $f(\boldsymbol{x})$ は $\boldsymbol{x}$ において極小になります．
- これは，最適性の 2 次の十分条件と呼ばれます．
- 以上のように，1 変数関数における最適性の条件が，多変数関数に拡張されます．

最急降下法



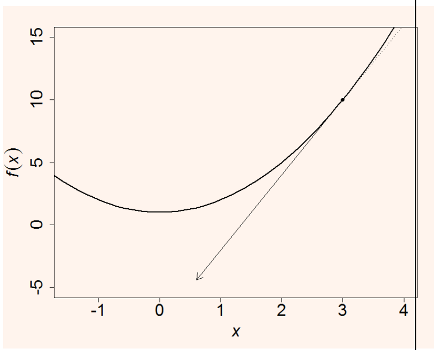
- ここでは，制約のない非線形最適化問題を解く方法の中で，最も基本的な方法である最急降下法について説明します．

最急降下法

- $f'(x)$ … 関数 $f(x)$ の増加の割合
- x_0 において

  $f'(x_0) > 0$　x を減少
  $f'(x_0) < 0$　x を増加

  … $f(x)$ を減少



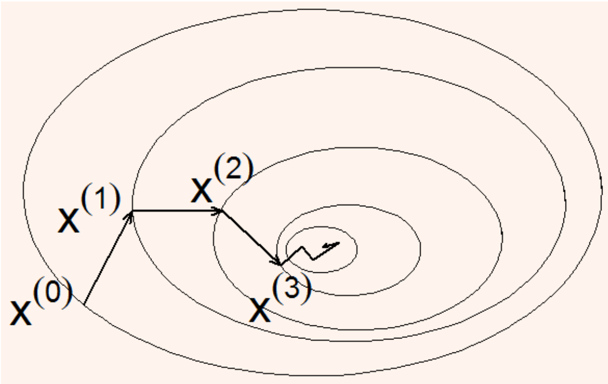
- まず，1 変数関数について考えてみましょう
- 関数 $f(x)$ の導関数，すなわち微分係数 $f'(x)$ は x における関数 $f(x)$ の増加の割合を示します．
- ですから，$x = x_0$ において，$f'(x_0)$ が正であれば x を十分小さく増加させれば，$f(x)$ の値は増加します．
- 逆に x を十分小さく減少させれば，$f(x)$ の値は減少します．
- 同様に，$f'(x_0)$ が負であれば x を十分小さく増加させれば，$f(x)$ の値は減少します．
- つまり，$x = x_0$ から，$f'(x_0)$ の値と逆方向に x を十分小さく変化させれば，$f(x)$ の値を小さくできますので，これを繰り返せば，$f(x)$ の極小値にたどり着くことができます．
- 次に，多変数関数について考えましょう．

最急降下法

- 勾配ベクトル $\nabla f(\boldsymbol{x})$ の向き
  $\cdots$ 関数 $f(\boldsymbol{x})$ を最も増加させる方向
- 勾配ベクトルの反対方向に $\boldsymbol{x}$ を変化させる
  $\cdots$ $f(\boldsymbol{x})$ を減少
- これを繰り返すことにより局所最適解に到達
  $\cdots$ **最急降下法**

- 勾配ベクトルの向きは関数 $f(\boldsymbol{x})$ を最も増加させる方向になりますので，$\boldsymbol{x}$ の各成分を勾配ベクトルの方向に十分小さく変化させると $f(\boldsymbol{x})$ の値は増加します．
- 逆に，$\boldsymbol{x}$ を勾配ベクトルの反対方向に十分小さく変化させると $f(\boldsymbol{x})$ は減少します．
- これを繰り返すことにより局所最適解に到達しようというのが最急降下法です．
- つまり，最急降下法は局所最適解を求める方法です．

最急降下法



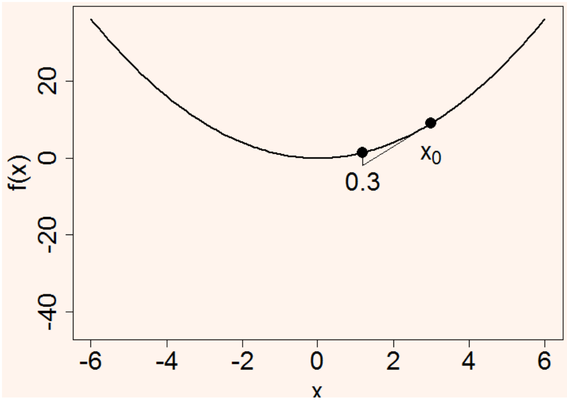
- 最急降下法のイメージを示します．
- 図は 2 変数関数の等高線で，中心部にいくほど値が小さくなるとします．
- 勾配ベクトルと等高線は直交します．
- $\boldsymbol{x}^{(0)}$ を初期値として，勾配ベクトルの逆方向に $\boldsymbol{x}^{(0)}$ を変化させ，変化後の値を $\boldsymbol{x}^{(1)}$ とします．
- $\boldsymbol{x}^{(1)}$ において勾配ベクトルを求めて，勾配ベクトルの逆方向に $\boldsymbol{x}^{(1)}$ を変化させます．
- これを繰り返すことにより，関数が極小になる点に到達します．

最急降下法

- k 回目の繰り返しにおいて $\boldsymbol{x}^{(k)}$ であるとき
$$\boldsymbol{x}^{(k+1)} \leftarrow \boldsymbol{x}^{(k)} - \alpha^{(k)} \nabla f(\boldsymbol{x}^{(k)})$$
と更新
- $\alpha^{(k)}$ は正の数
$$f(\boldsymbol{x}^{(k)} - \alpha^{(k)} \nabla f(\boldsymbol{x}^{(k)}))$$
が極小に (近く) なるように定める

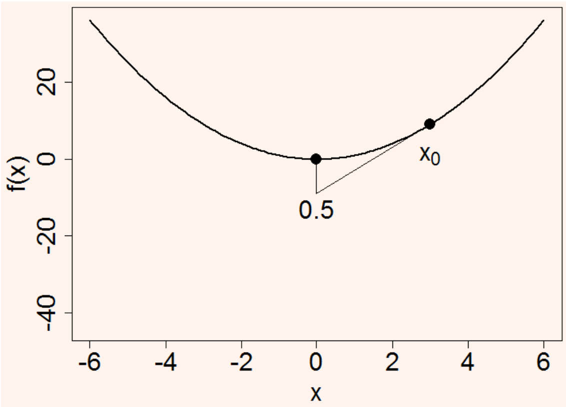
- 最急降下法では，k 回目の繰り返しにおいて $\boldsymbol{x}^{(k)}$ であるとき，この式により $\boldsymbol{x}^{(k)}$ を更新します．
- ここで，$\alpha^{(k)}$ は正の数で，更新後の関数の値が極小になるように，現実的には極小に近くなるように定めます．

最急降下法



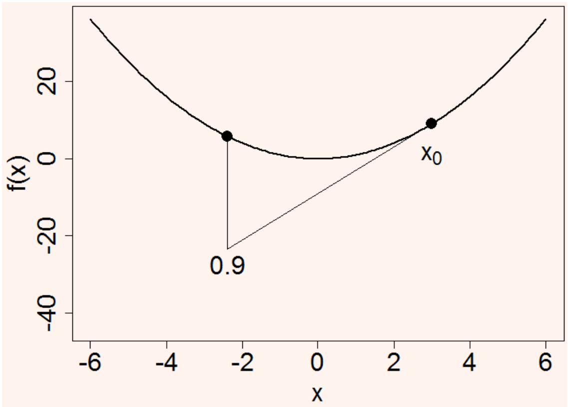
- 勾配ベクトルの逆向きは関数を減少させる方向ですから，$\alpha^{(k)}$ が十分小さければ，関数の値は必ず減少します．
- しかし，$\alpha^{(k)}$ が小さすぎると，関数はあまり減少しません．
- この図は x_0 から $\alpha^{(0)} = 0.3$ で更新した場合です．

最急降下法



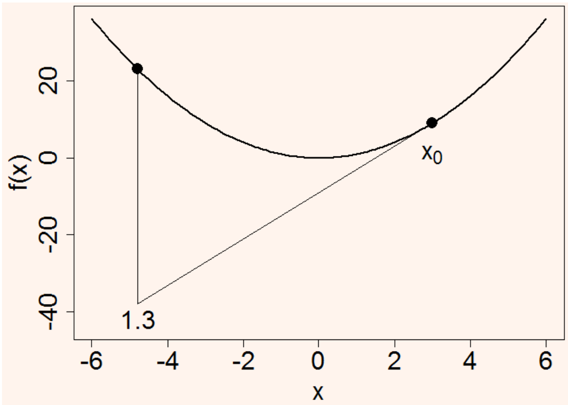
- 今度は $\alpha^{(0)} = 0.5$ で更新した場合です.
- この時，関数は極小値に近くなっています.

最急降下法

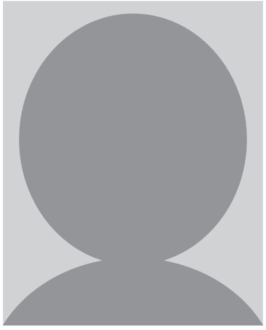


- 今度は $\alpha^{(0)} = 0.9$ で更新した場合です．
- この場合，関数が極小になる点を超えていて，元の x_0 よりは関数の値は小さくなっていますが，最適からはかなりずれています．
- この場合，$\alpha^{(0)}$ は最適値より大き過ぎます．

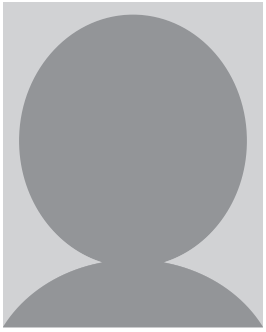
最急降下法



- 今度は $\alpha^{(0)} = 1.3$ で更新した場合です．
- この場合，関数が極小になる点をはるかに超えて，元の x_0 より関数の値は大きくなってしまいます．
- この場合，$\alpha^{(0)}$ は最適値より大き過ぎます．
- このように，$\alpha^{(0)}$ は小さすぎても，大きすぎてもよろしくありません．



- 先ほどお話しましたように，$\alpha^{(k)}$ が十分小さければ，関数の値は必ず減少します．
- もし，暫定的に定めた $\alpha^{(k)}$ では関数が増加する場合，$\alpha^{(k)}$ は大きすぎますから，関数が減少するまで繰り返し $\alpha^{(k)}$ を小さくしていきます．
- もし，暫定的に定めた $\alpha^{(k)}$ で関数が減少する場合，関数が極小になる点を超えるまで繰り返し $\alpha^{(k)}$ を大きくしていきます．



- $\alpha^{(k)}$ の変化の幅を小さくしすぎると，$\alpha^{(k)}$ の決定に時間がかかりますので，そういう場合は，現在の値の 2 倍とか 1/2 倍くらいの幅で $\alpha^{(k)}$ を変化させて，適当な方法で補間をして，関数の値が極小に近くなる $\alpha^{(k)}$ を推定します．
- その推定法は幾つか提案されていますが，印刷教材には黄金分割法と呼ばれる手法を載せてありますので，具体的な手続きは印刷教材をご覧ください．

最急降下法

(0) $\boldsymbol{x}$ の適当な初期値 $\boldsymbol{x}^{(0)}$ を定める　$k \leftarrow 0$

(1) $\nabla f(\boldsymbol{x}^{(k)}) = \boldsymbol{0}$ ならば $\boldsymbol{x}^{(k)}$ を局所最適解として終了
そうでなければ (2) へ

(2) $f(\boldsymbol{x}^{(k)} - \alpha^{(k)} \nabla f(\boldsymbol{x}^{(k)}))$ が極小に (近く) なるような
$\alpha^{(k)}$ を求め

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha^{(k)} \nabla f(\boldsymbol{x}^{(k)})$$

により $\boldsymbol{x}^{(k)}$ を更新　$k \leftarrow k+1$ として (1) へ

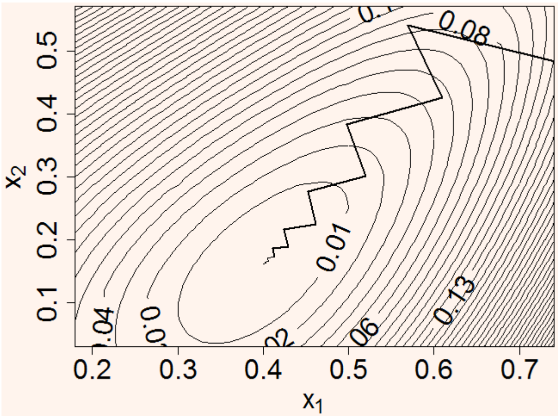
- 最急降下法のアルゴリズムを示します．
- まず，(0) で初期化を行います．
- $\boldsymbol{x}$ の適当な初期値 $\boldsymbol{x}^{(0)}$ を定めます．
- また，繰り返し回数のカウンター k を 0 とします．
- 次に (1) では，勾配ベクトルが零ベクトルであれば，$\boldsymbol{x}^{(k)}$ を局所最適解として，計算を終了します．
- そうでなければ，(2) へ進みます．
- (2) では，勾配ベクトルの逆方向に $\boldsymbol{x}^{(k)}$ を更新します．
- 変化の幅 $\alpha^{(k)}$ は関数が極小値の近くになるように適当な方法で定めます．
- そして，k の値を 1 増やして，以降，(1), (2) を繰り返します．

最急降下法

- 実際の終了基準
  小さな正の数 ε に対して
  $|\nabla f(\boldsymbol{x}^{(k)})| < \varepsilon$
- $\boldsymbol{x}$ の任意の初期値に対して局所最適解に収束する
  … **大域収束性**
- 収束までに多数の繰り返し数を要することが多い

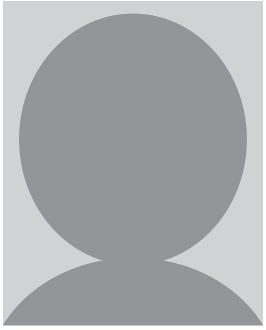
- 終了基準は勾配ベクトルが零ベクトルとしましたが，それではなかなか終了しませんので，実際には，例えば勾配ベクトルのノルムが十分小さいといったことを終了条件にします．
- ベクトルのノルムとはベクトルの原点からの距離のことですが，要するにベクトルの各成分の絶定値が十分 0 に近ければ終了します．
- 最急降下法は，$\boldsymbol{x}$ の任意の初期値に対して局所最適解に収束することが保証されています．
- この性質を大域収束性といいます．
- しかし，収束までに多数の繰り返しを要することが多く，効率は次に説明するニュートン法に劣ります．

最急降下法



- これは最急降下法で 2 変数関数の最小化を行った例ですが，等高線の中心部，すなわち局所最適解に近づくと小刻みに $\boldsymbol{x}$ が更新されています．
- 後で，ニュートン法と比較しますが，収束までに多数の繰り返しを要します．

ニュートン法



- ここではニュートン法について説明します.
- ニュートン法は収束までの繰り返しが圧倒的に少ないという特徴を持っています.

ニュートン法

- 1 変数関数 $f(x)$ は，定数 $x^{(k)}$ のまわりで
テイラー展開により

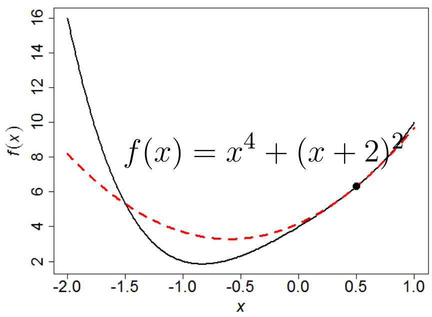
$$f(x) = f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)}) + \frac{1}{2!}f''(x^{(k)})\left(x - x^{(k)}\right)^2 + \frac{1}{3!}f'''(x^{(k)})\left(x - x^{(k)}\right)^3 + \cdots$$

と近似できる

- 1 次関数，2 次関数で近似することが多い

- まず，1 変数関数で考えます.
- 変数 x が定数 $x^{(k)}$ の近くにある時，関数 $f(x)$ はここに示すような x の多項式で近似できます.
- これはテイラー展開と呼ばれます.
- 多くの場合，計算を簡単にするため，1 次関数や 2 次関数で近似します.

ニュートン法



$$g(x) = f(0.5) + f'(0.5)(x - 0.5) + \frac{1}{2}f''(0.5)(x - 0.5)^2$$

- 一つ例を示します．
- $f(x)$ はグラフ内に示す 4 次関数で，黒い実線でプロットされています．
- これを $x^{(k)} = 0.5$ の周りのテイラー展開により 2 次式で近似したものが下の式で表される $g(x)$ です．
- $g(x)$ は赤い破線でプロットされています．
- $x = 0.5$ 付近では，$g(x)$ はかなり精度よく $f(x)$ を近似できていることが分かります．
- もちろん，x が 0.5 から離れると近似精度は落ちます．

ニュートン法

- $f(x)$ を $x^{(k)}$ のまわりで2次関数$g(x)$ と近似する

$$g(x) = \begin{aligned} & f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)}) \\ & + \frac{1}{2} f''(x^{(k)})(x - x^{(k)})^2 \end{aligned}$$

- $f''(x^{(k)}) > 0$ と仮定すれば
最適性の 1 次の必要条件

$$g'(x) = 0$$

を満たす x において $g(x)$ は最小

- さて，1 変数関数 $f(x)$ を定数 $x^{(k)}$ の周りでテイラー展開した 2 次関数 $g(x)$ で近似したとしましょう．
- $x = x^{(k)}$ において，$f(x)$ の 2 階微分の値が正であると仮定すると，$g(x)$ の 1 階微分である $g'(x)$ が 0 になる x において，$g(x)$ は最小になります．
- $g(x)$ は 2 次関数ですから，極小であるだけでなく最小になります．

ニュートン法

- 最適解 x は

$$g'(x) = f'(x^{(k)}) + f''(x^{(k)})(x - x^{(k)}) = 0$$

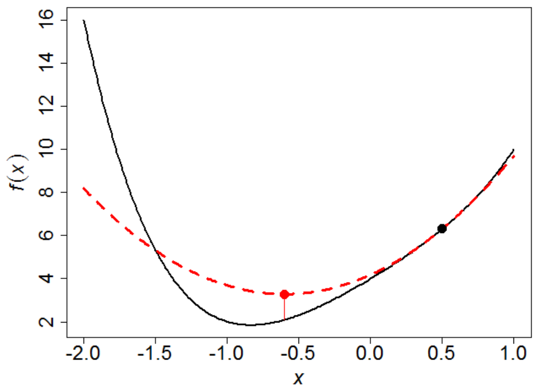
を解いて

$$x = x^{(k)} - f'(x^{(k)})/f''(x^{(k)})$$

- x は $f(x)$ の近似式から導いたので近似解
- 更新を繰り返して $f(x)$ の局所最適解に到達
… **ニュートン法**

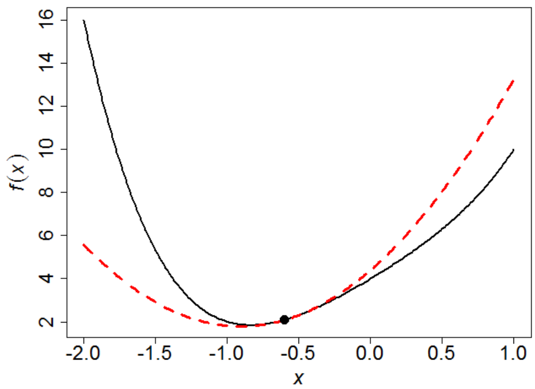
- 方程式 $g'(x) = 0$ は上の式のようになります．
- この方程式は簡単に解けまして，下の式のようになります．
- ただし，この x は関数 $f(x)$ の近似式 $g(x)$ から導いたものですので，関数 $f(x)$ の極小値を与える x としては近似解です．
- そこで，この近似解の周りで関数 $f(x)$ を 2 次式で近似し，解の更新を繰り返して，元の関数 $f(x)$ の局所最適解に到達しようというのがニュートン法です．

ニュートン法



- 先ほどの 4 次関数 $f(x)$ の例では，$g(x)$ の最小値を与える x は -0.6 です．
- グラフを見れば，$f(x)$ の極小値を与える x とはまだ差があることが分かります．

ニュートン法



- そこで，$x = -0.6$ の周りで，$f(x)$ をテイラー展開して，2 次関数で近似し，その最小値を与える x を求めます．
- グラフから，2 次関数を最小にする x は，$f(x)$ を最小にする x とかなり近くなることが分かります．
- このように 2 次関数での近似とその最適解を求めることを繰り返すことにより，元の関数の局所最適解が得られます．

ニュートン法

- 多変数関数 $f(\boldsymbol{x})$ は，定数ベクトル $\boldsymbol{x}^{(k)}$ のまわりでテイラー展開により

$$
\begin{aligned}
f(\boldsymbol{x}) \simeq g(\boldsymbol{x}) = \quad & f(\boldsymbol{x}^{(k)}) + \nabla f(\boldsymbol{x}^{(k)})^T(\boldsymbol{x} - \boldsymbol{x}^{(k)}) \\
& + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^{(k)})^T \nabla^2 f(\boldsymbol{x}^{(k)})(\boldsymbol{x} - \boldsymbol{x}^{(k)}) \\
& + \cdots
\end{aligned}
$$

と近似できる

- 今度は，多変数関数で考えてみましょう.
- 多変数関数 $f(\boldsymbol{x})$ は点 $\boldsymbol{x}^{(k)}$ のまわりで，この式のように近似できます.
- この近似した関数を $g(\boldsymbol{x})$ とします.
- $g(\boldsymbol{x})$ は，テイラー展開の多変数版ということになります.

ニュートン法

- $f(\boldsymbol{x})$ を $\boldsymbol{x}^{(k)}$ のまわりで2次関数$g(\boldsymbol{x})$ と近似する

$$g(\boldsymbol{x}) = \quad f(\boldsymbol{x}^{(k)}) + \nabla f(\boldsymbol{x}^{(k)})^T(\boldsymbol{x} - \boldsymbol{x}^{(k)}) \\ +\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^{(k)})^T \nabla^2 f(\boldsymbol{x}^{(k)})(\boldsymbol{x} - \boldsymbol{x}^{(k)})$$

- ヘッセ行列 $\nabla^2 f(\boldsymbol{x}^{(k)})$ が正定値と仮定すれば
  最適性の 1 次の必要条件

$$\nabla g(\boldsymbol{x}) = \boldsymbol{0}$$

  を満たす $\boldsymbol{x}$ において $g(\boldsymbol{x})$ は最小

- 関数 $f(\boldsymbol{x})$ を $\boldsymbol{x}^{(k)}$ のまわりで 2 次関数 $g(\boldsymbol{x})$ で近似したとします．
- ここで，ヘッセ行列が正定値であると仮定します．
- 1 変数関数では，2 階微分の値が正であることを仮定しましたが，多変数関数ではヘッセ行列が正定値であると仮定します．
- そうしますと，勾配ベクトルが零ベクトルであれば，最適性の 2 次の十分条件を満たしますので，$g(\boldsymbol{x})$ は最小となります．
- そのため，$g(\boldsymbol{x})$ の勾配ベクトルが零ベクトルになる $\boldsymbol{x}$ を求めます．
- ここで，$x^{(k)}$ は各成分の値が決まっていますから，$f(\boldsymbol{x}^{(k)})$ は定数，それから $x^{(k)}$ における勾配ベクトルやヘッセ行列は定数ベクトル，定数行列です．
- $g(\boldsymbol{x})$ は変数 $\boldsymbol{x}$ の 2 次関数です．
- 以上のことを注意して $g(\boldsymbol{x})$ の勾配ベクトルを求めると，下のような式になります．
- 勾配ベクトルが零ベクトルであるという方程式を解きますと，$g(\boldsymbol{x})$ を最小化する $\boldsymbol{x}$ を求めることができます．

ニュートン法

- 最適解 $\boldsymbol{x}$ は

$$\nabla g(\boldsymbol{x}) = \nabla f(\boldsymbol{x}^{(k)}) + \nabla^2 f(\boldsymbol{x}^{(k)})(\boldsymbol{x} - \boldsymbol{x}^{(k)}) = \mathbf{0}$$

を解いて

$$\boldsymbol{x} = \boldsymbol{x}^{(k)} - \nabla^2 f(\boldsymbol{x}^{(k)})^{-1} \nabla f(\boldsymbol{x}^{(k)})$$

- $\boldsymbol{x}$ は $f(\boldsymbol{x})$ の近似式から導いたので近似解
- 更新を繰り返して $f(\boldsymbol{x})$ の局所最適解に到達
… **ニュートン法**

- この方程式は簡単に解けまして，下の式のようになります．
- ただし，$\boldsymbol{x}$ は $f(\boldsymbol{x})$ の近似式から導いたものですので，近似解です．
- そこで，この近似解の周りで，$f(\boldsymbol{x})$ を 2 次関数で近似して，その最適解を求めて，元の関数 $f(\boldsymbol{x})$ の近似解の更新を繰り返して，元の関数 $f(\boldsymbol{x})$ の局所最適解に到達しようというのがニュートン法です．

ニュートン法

(0) $\boldsymbol{x}$ の適当な初期値 $\boldsymbol{x}^{(0)}$ を定める　$k \leftarrow 0$

(1) $\nabla f(\boldsymbol{x}^{(k)}) = \mathbf{0}$ ならば
$\boldsymbol{x}^{(k)}$ を局所最適解として計算終了
そうでなければ (2) へ

(2) $\boldsymbol{x}^{(k+1)} \leftarrow \boldsymbol{x}^{(k)} - \nabla^2 f(\boldsymbol{x}^{(k)})^{-1} \nabla f(\boldsymbol{x}^{(k)})$
により $\boldsymbol{x}^{(k)}$ を更新　$k \leftarrow k+1$ として，(1) へ

- ニュートン法のアルゴリズムを示します．
- まず，(0) で初期化を行います．
- $\boldsymbol{x}$ の適当な初期値 $\boldsymbol{x}^{(0)}$ を定めます．
- また，繰り返し回数のカウンター k を 0 とします．
- 次に (1) では，勾配ベクトルが零ベクトルであれば，$\boldsymbol{x}^{(k)}$ を局所最適解として，計算を終了します．
- そうでなければ，(2) へ進みます．
- (2) では，ヘッセ行列の逆行列と勾配ベクトルの積，これはベクトルになりますが，このベクトルの逆方向に $\boldsymbol{x}^{(k)}$ を更新します．
- そして，k の値を 1 増やして，以降，(1), (2) を繰り返します．
- ニュートン法の特徴は局所最適解への収束が極めて速いことです．
- 数値例を見てみましょう．

ニュートン法

数値例

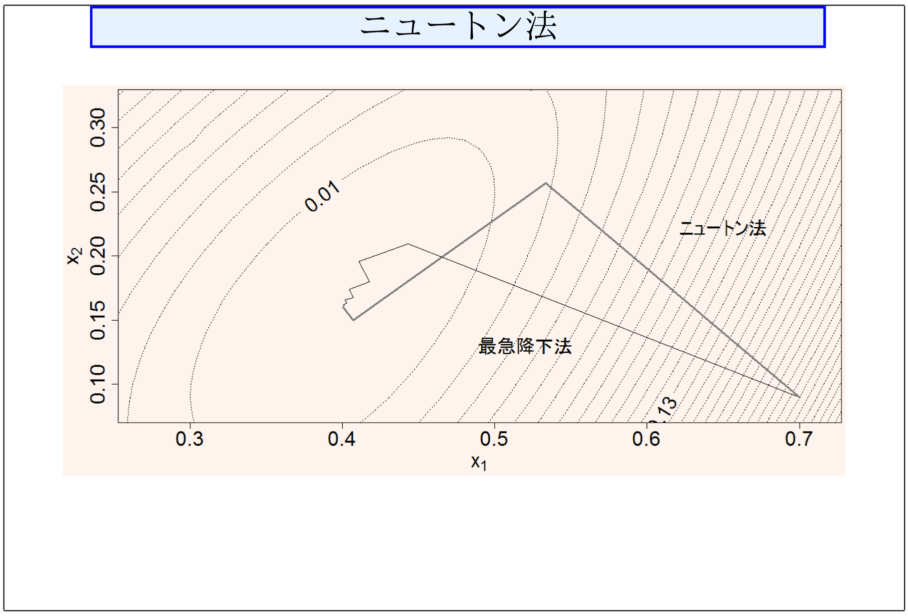
$$f(\boldsymbol{x}) = (x_1 - 0.4)^2 + (x_1^2 - x_2)^2$$

を最小化する $\boldsymbol{x}$ を最急降下法とニュートン法を用いて求める

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2(x_1 - 0.4) + 4x_1(x_1^2 - x_2) \\ -2(x_1^2 - x_2) \end{bmatrix},$$

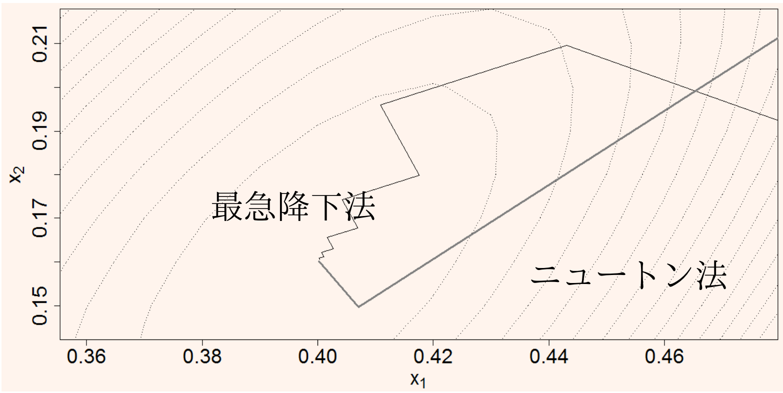
$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 12x_1^2 - 4x_2 + 2 & -4x_1 \\ -4x_1 & 2 \end{bmatrix}$$

- 2 変数関数 $f(x) = (x_1 - 0.4)$ の 2 乗＋ $(x_1^2 - x_2)$ の 2 乗を最小化する x_1, x_2 を求めます.
- 最適解は，$x_1 = 0.4$, $x_2 = x_1^2 = 0.16$ であることは明らかですが，これを最急降下法とニュートン法で計算します.
- 勾配ベクトルとヘッセ行列はここに示す通りです.
- 初期値は $x_1 = 0.7$, $x_2 = 0.1$ として，勾配ベクトルの成分の絶対値の最大値が 10 の -3 乗より小さくなることを終了条件としました.
- また，最急降下法における，更新の幅 $\alpha^{(k)}$ は黄金分割法で決定しました.



- 計算結果はこの図の通りで，ニュートン法は 3 回の更新で収束しています．
- 一方，最急降下法では 10 回の更新で収束しています．
- 局所最適解付近を拡大しますと，次のようになります．

ニュートン法

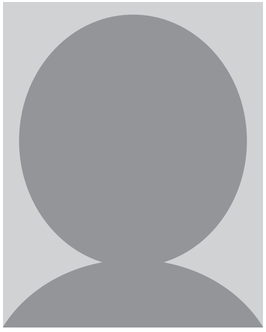


- 最急降下法では，ちまちまとした更新を繰り返すのに対して，ニュートン法では一気に局所最適解に迫っていることが分かります．

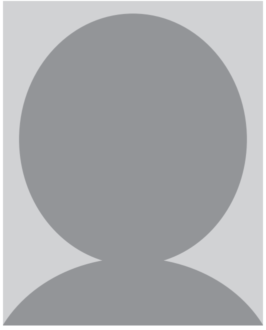
ニュートン法

- 局所最適解への収束が極めて速い
- ヘッセ行列が常に正定値であることを仮定
  これが満たされない場合
  → 局所最適解への収束が保証されない
- 初期値が局所最適解に十分近ければ，
  局所最適解への収束が保証される… 局所収束性

- このように，ニュートン法は局所最適解へ極めて速く収束するという特徴を持っていますが，欠点もあります．ニュートン法はヘッセ行列が常に正定値であることを仮定していますが，これが満たされない時には局所最適解への収束が保証されません．
- 局所最適解においてヘッセ行列は正定値になりますから，初期値が局所最適解に十分近ければ，ニュートン法で局所最適解に収束することが保証されます．
- この性質を局所収束性と呼びます．



- ニュートン法は収束がきわめて速く，魅力的な方法ですが，大域収束性がないのは辛いところです.
- この問題の対策は幾つかあります.
- まず，最初は最急降下法で計算を行い，ある程度更新を重ねた後にニュートン法に切り替えるという方法です
- この方法は簡単で，実用的であると思いますが，どの時点でニュートン法に切り替えるかを決めるには試行錯誤が必要かもしれません.



- 理論的にはもっとエレガントな方法もありますが，計算が複雑になります.
- もう一つの方法は,「準ニュートン法」と呼ばれる方法です.

ニュートン法

準ニュートン法

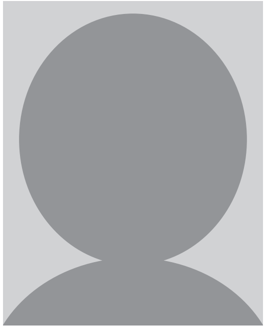
- ヘッセ行列を勾配ベクトルなどから逐次近似

BFGS 法

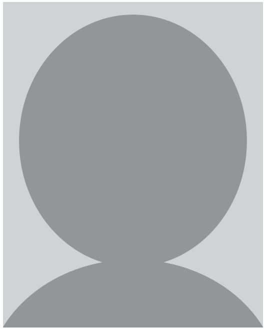
$$\boldsymbol{B}^{(k+1)} = \boldsymbol{B}^{(k)} + \frac{\boldsymbol{y}^{(k)}\boldsymbol{y}^{(k)T}}{\boldsymbol{y}^{(k)T}\boldsymbol{s}^{(k)}} - \frac{\boldsymbol{B}^{(k)}\boldsymbol{s}^{(k)}\boldsymbol{s}^{(k)T}\boldsymbol{B}^{(k)T}}{\boldsymbol{s}^{(k)T}\boldsymbol{B}^{(k)}\boldsymbol{s}^{(k)}}$$

$$\boldsymbol{B}^{(0)} = \boldsymbol{I},\ \boldsymbol{s}^{(k)} = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)},\ \boldsymbol{y}^{(k)} = \nabla f(\boldsymbol{x}^{(k+1)}) - \nabla f(\boldsymbol{x}^{(k)})$$

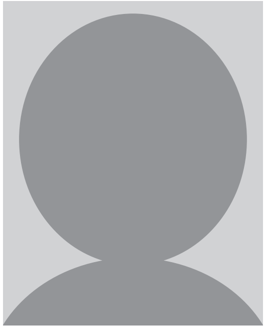
- ニュートン法は，ヘッセ行列が正定値にならないことがあるだけでなく，変数が多い時には，ヘッセ行列の計算に時間がかかったりするという欠点があります．
- 準ニュートン法はヘッセ行列を直接計算せず，勾配ベクトルなどからヘッセ行列を逐次近似する手法です．
- ヘッセ行列の近似法はいろいろ提案されていますが，中でもよく知られているのは BFGS 法です．
- 参考までに BFGS 法によるヘッセ行列の近似式を示します．
- 複雑に見えるかもしれませんが，行列の基本演算だけなので，それほど難しくはありません．
- 準ニュートン法は，面倒なヘッセ行列の計算がなく，収束が速いという便利さを備えています．
- さらに，特に条件を加えなくても，実際，大域収束性を持つ場合が多く，この点でも優れていると言えます．
- また，多少の修正を加えれば大域収束性が保証されます．



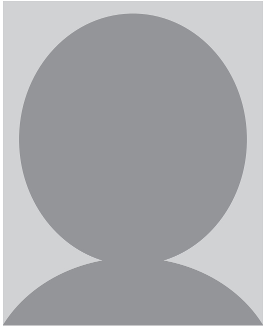
- 15-20 秒版
- 今回の講義では，非線形最適化問題の定式化と，制約のない非線形最適化問題の解法についてお話しました．
- 内容が盛りだくさんでしたので，よく復習して，基本的な考え方の理解に努めて下さい．
- 今回はこれで終わります．



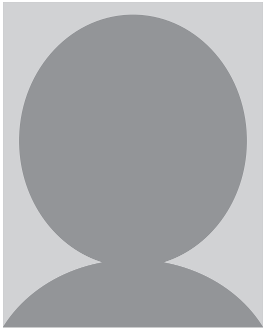
- 15-20 秒版
- 今回の講義では，非線形最適化法の基礎として，非線形最適化問題の定式化と，制約のない非線形最適化問題の解法についてお話しました．
- 内容が盛りだくさんで，今回は難しく感じられたかもしれません．
- よく復習して，基本的な考え方の理解に努めるようにして下さい．
- 時間が来ましたので，今回はこれで終わります．



- 45 秒版
- 今回の講義では，非線形最適化問題の定式化と，制約のない非線形最適化問題の解法についてお話しました．
- 今回は，テイラー展開や偏微分，勾配ベクトルにヘッセ行列，さらには半正定値と正定値と，あまりお馴染みではなさそうな概念が満載でしたので，難しかったかもしれません．
- しかし，基本的な考え方は，関数の極値を与える点，一階微分の値がゼロになる点を求めるという，高校数学でも習う微分の応用の拡張ですので，まずは基本的な考え方の理解に努めて下さい．
- 今回はこれで終わります．



- 45 秒版
- 今回の講義では，非線形最適化問題の定式化と，制約のない非線形最適化問題の解法についてお話しました．
- 今回は，テイラー展開や偏微分，勾配ベクトルにヘッセ行列，さらには半正定値と正定値と，あまりお馴染みではなさそうな概念が満載でしたので，難しかったかもしれません．
- しかし，基本的な考え方は，関数の極値を与える点，一階微分の値がゼロになる点を求めるという，高校数学でも習う微分の応用の拡張ですので，まずは基本的な考え方の理解に努めて下さい．



- 今回は，解法の説明に時間を割きましたが，実際の計算は計算機ソフトウエアに任せることができますので，計算の詳細よりは問題の定式化に力を入れて復習していただきたいと思います.
- 今回はこれで終わります.